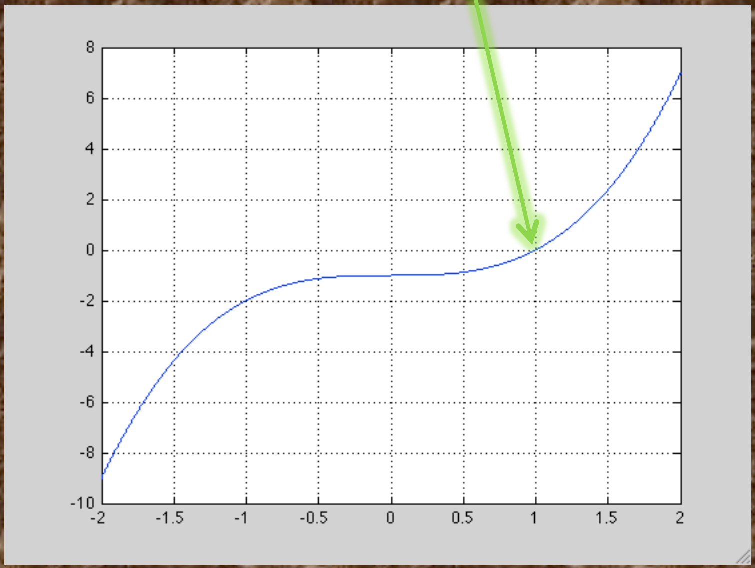
Newton's Method

MATLAB

Newton's method is a numerical technique for finding the (approximate) root(s) of the equation $f(x)=0$.

Below we see a graph of the equation $f(x)=x^3-1=0$, which has a real root at $x=1$.



The idea of Newton's method is to generate a new approximation to the solution by using the intercept of the tangent line at the current approximation. This process is repeated until an acceptable approximation to the root is obtained.

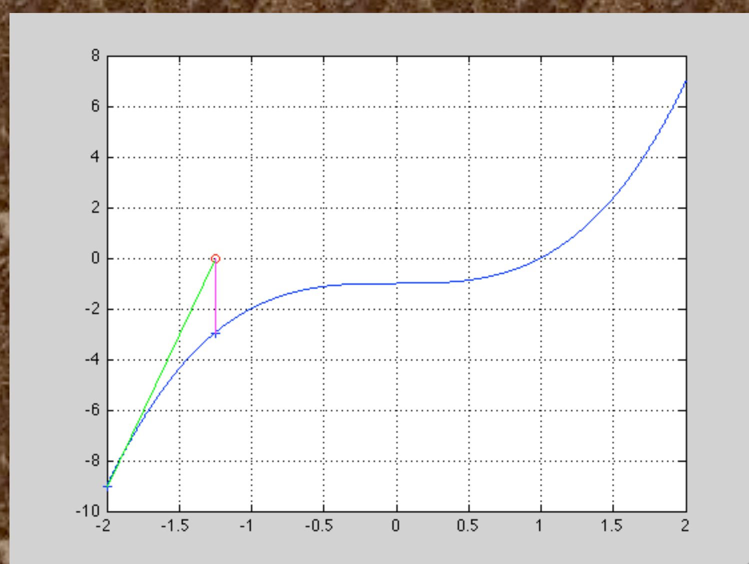
If we know the analytic form of the function we can use that to calculate the derivative (tangent).

If we do not know the analytic form of the function, we can calculate the derivative (tangent) from the data ($\Delta y/\Delta x$).

For the next approximation, we have $y=mx+b$, the equation for a line

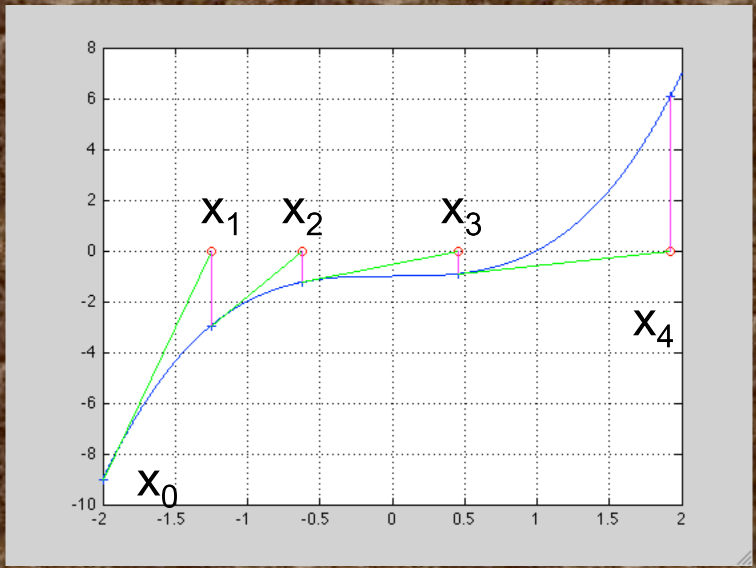
$$f(x_0)=y_0=mx_0+b$$
$$y_1=mx_1+b \quad \text{with } y_1=0$$
So $x_{n+1}=x_n+f(x_n)/f'(x_n)$ with $f'(x_n)\neq 0$.

Using this, for $x_0=-2$, $x_1\approx-1.25$

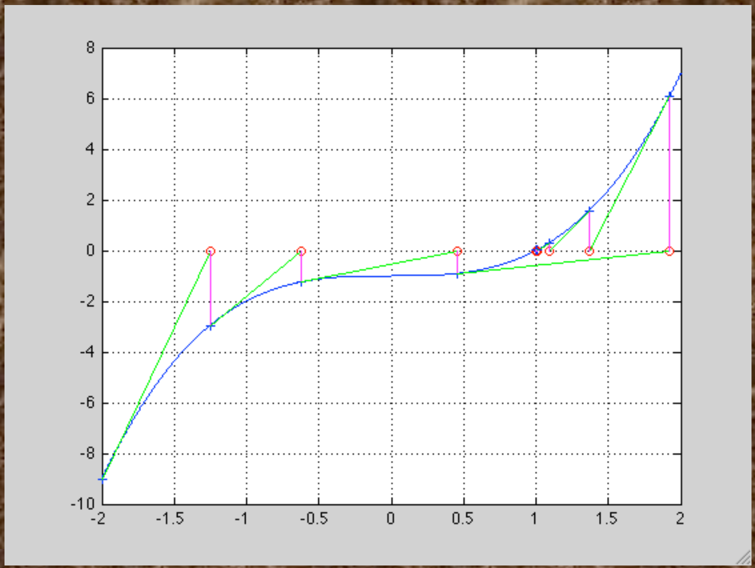


x_n approaches the root x=0 for the first few iterations, then diverges slightly. This is due to the small value of the slope at x_3.

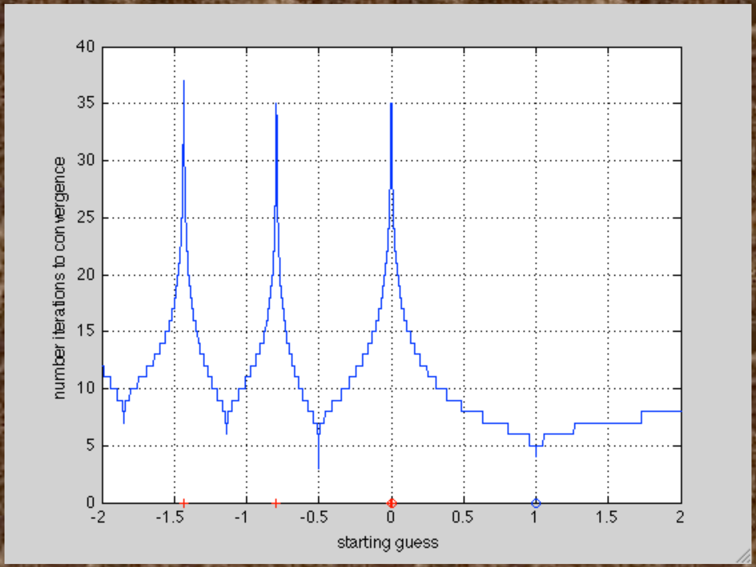
(and this shows the problem when f'=0, the new approximation shoots off to infinity. More on this later.)



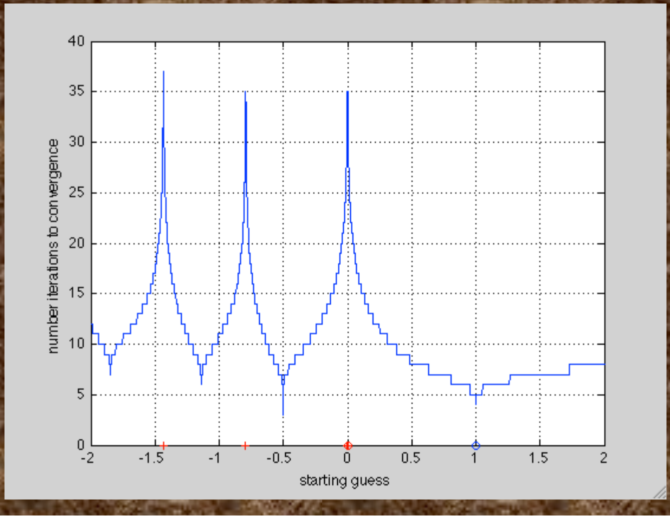
The 5th iteration, however, starts another converging sequence that visually converges after a total of only 8 iterations.



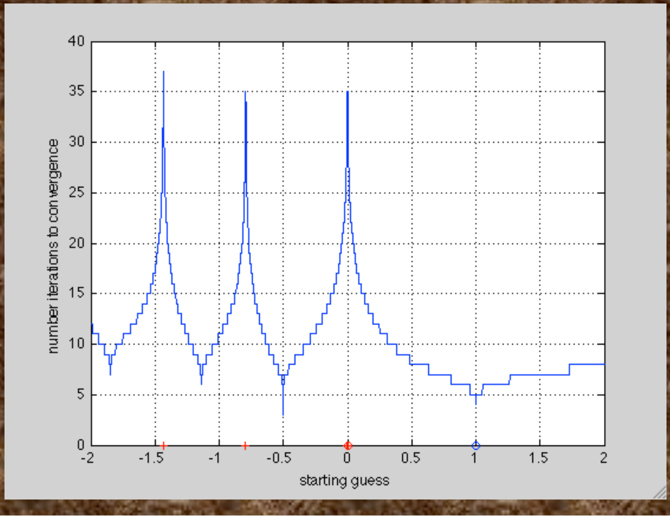
What happens if I chose another starting value? The plot below shows the number of iterations to arrive at the solution for all starting values between -2 and 2.



Notice there are several starting values for which convergence is not reached: the red circle at x=0 indicates the method fails there immediately due to the slope being zero, while the red "+" signs indicate convergence has not been reached after 30 iterations.

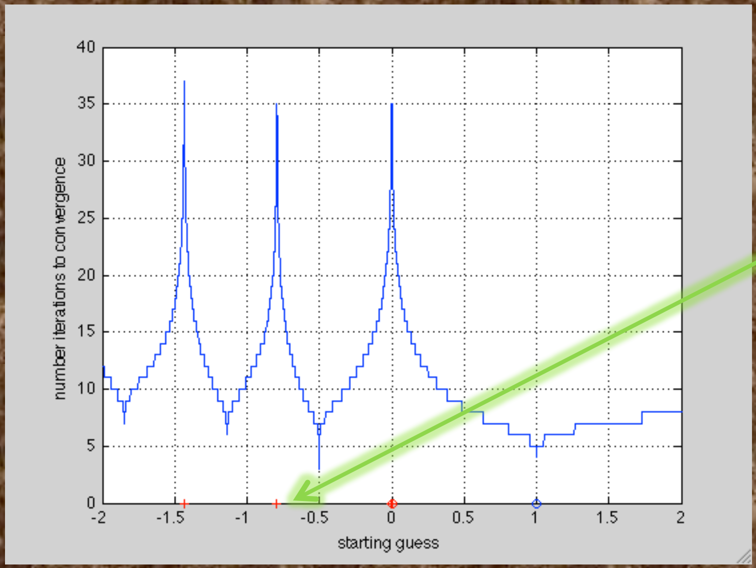


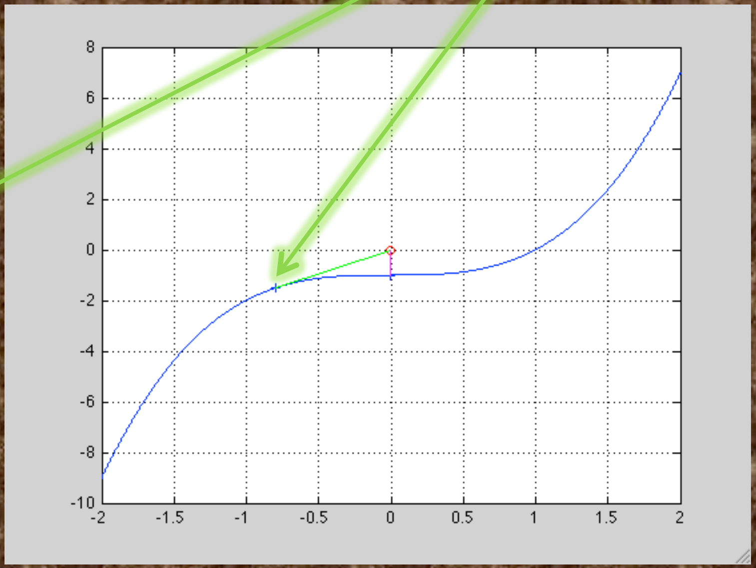
In general, for most starting values there is no simple way to predict the rate of convergence from the starting value.



We can easily find the first non-zero starting value that diverges - it is the one that produces x=0 at the next guess.

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad x_1 = 0 = x_0 - \frac{x_0^3}{3x_0^2}, \quad x_0 = \sqrt[3]{-0.5} = -0.7937\ldots$$

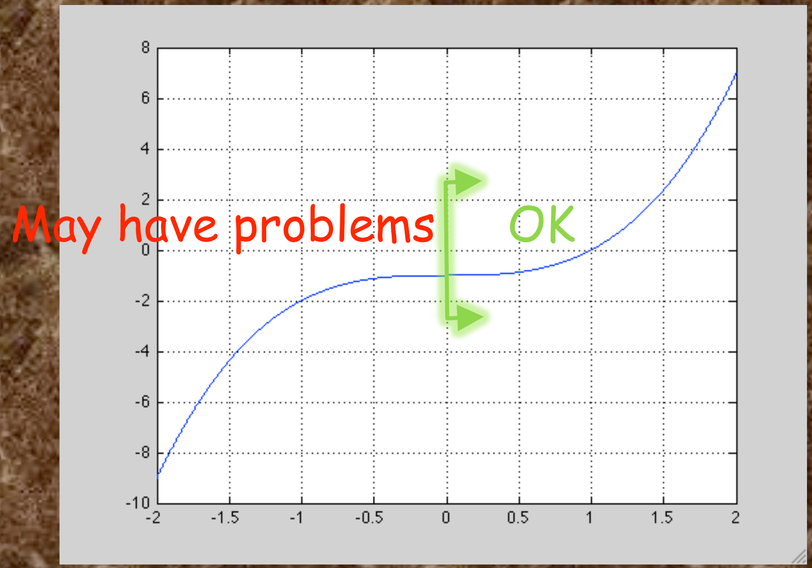




We can continue calculating the "next" bad starting value.

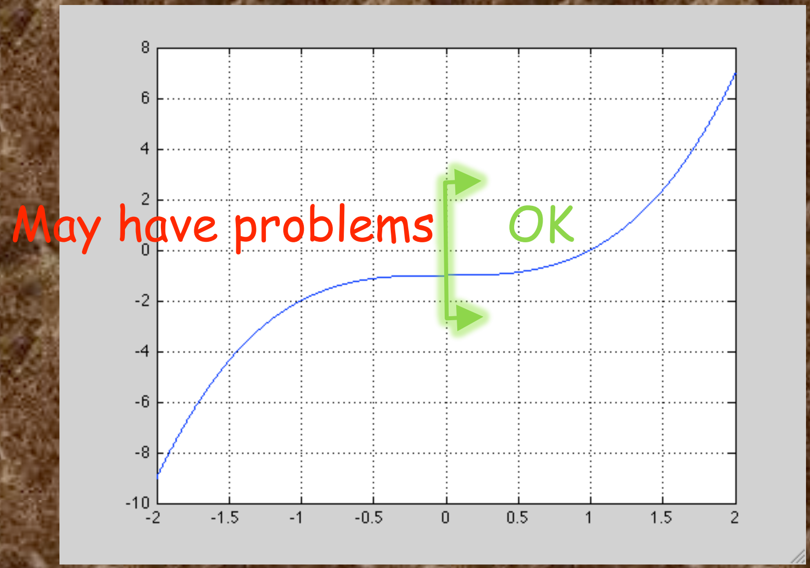
As long as there are no places where the derivative is zero between where you are and the solution, you are OK (you will converge).

If there is a place with zero slope between you and the solution, you will have problems if you get near that spot.



In this case, where we know the function and derivative, you can find the starting values that will have problems by working backwards.

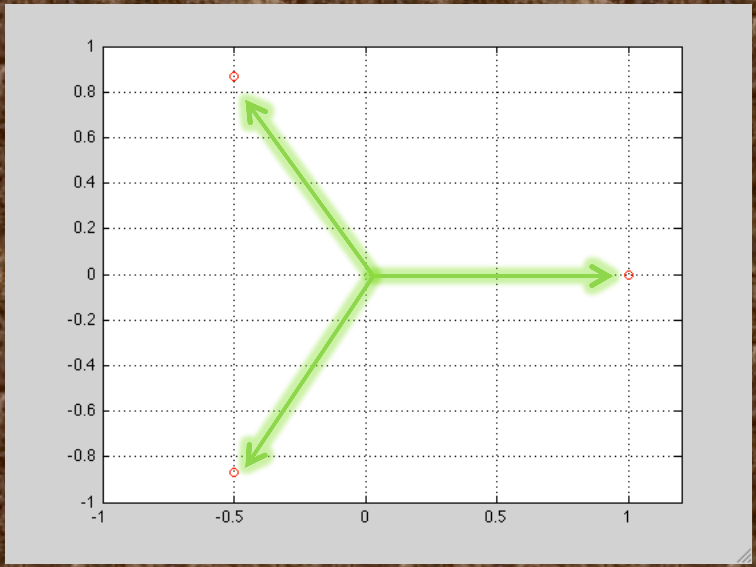
If you don't have f(x) analytically, it is harder to identify the bad starting values.



This is not the full solution however.

So far we have only considered the real part of the problem.

We know that $x^3-1=0$ has 3 solutions.
(magnitude =1, at 120° around the origin.)



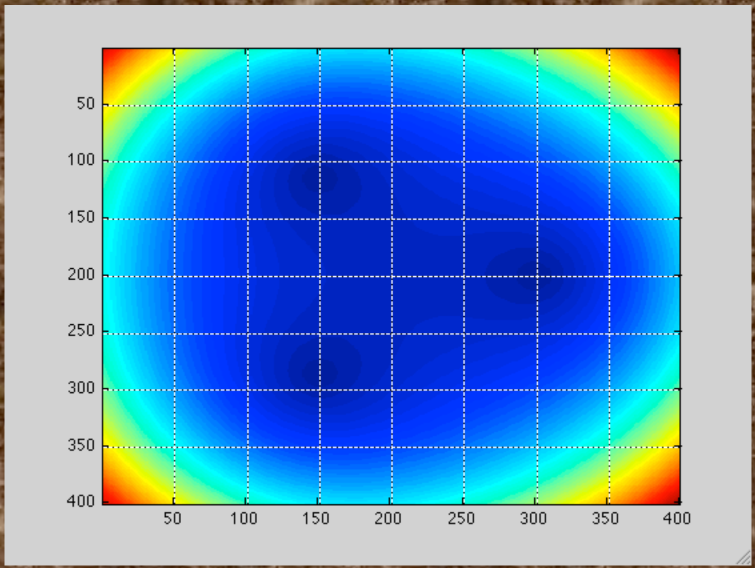
So now let's try Newton's method again, but now letting x be complex (and calling it z), and taking starting points in the 4x4 box around the origin as starting points.

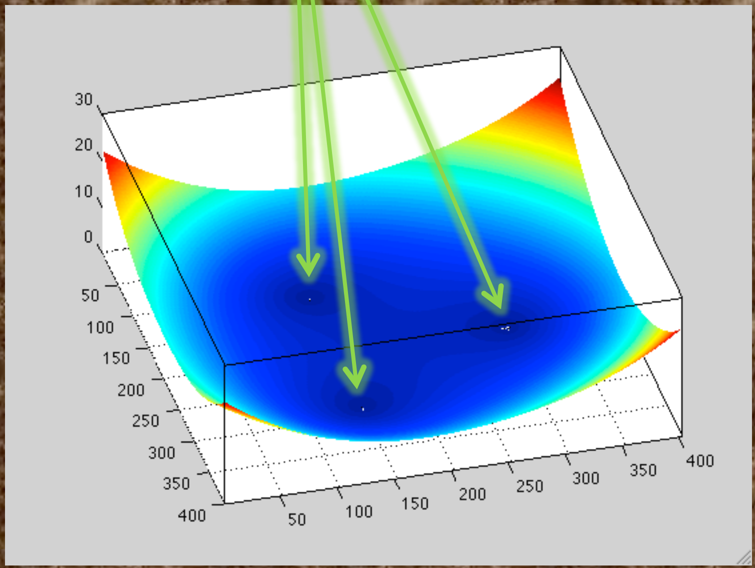
(many presentations of Newton's method say f has to be a real valued function, but this restriction is not needed. Just use the complex derivative, which is like the gradient – it will also have a direction.)

We will count how many iterations to convergence plus keep track of which root Newton's method converges to. We will indicate the root by color (so we will have 3 colors), and use brightness to indicate the speed of convergence (brighter is faster).

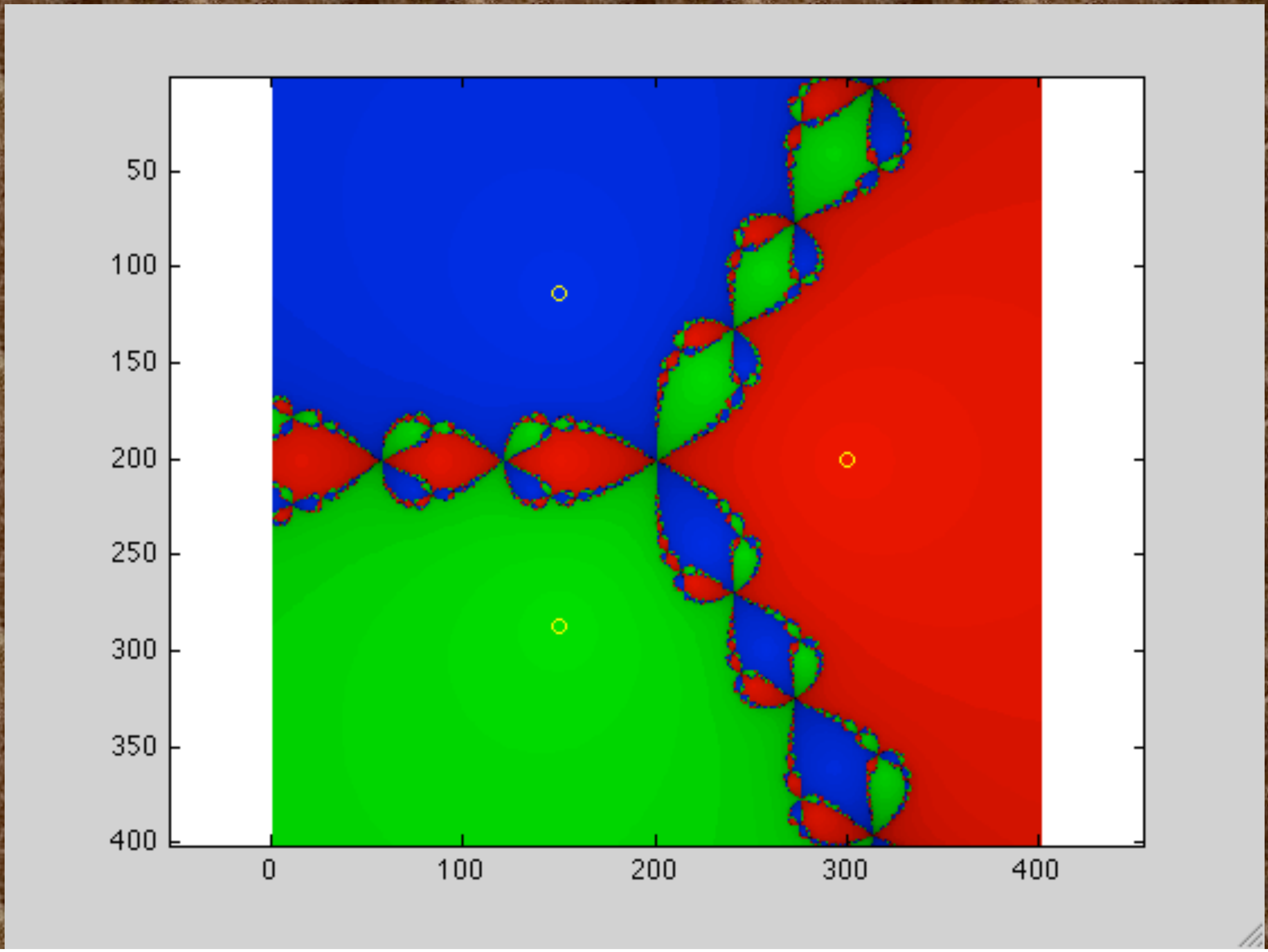
Here are some plots of the absolute value of $f(z)$ ($f(z)$ is a complex number so we can't plot it directly).

We are aiming for the 3 low spots.



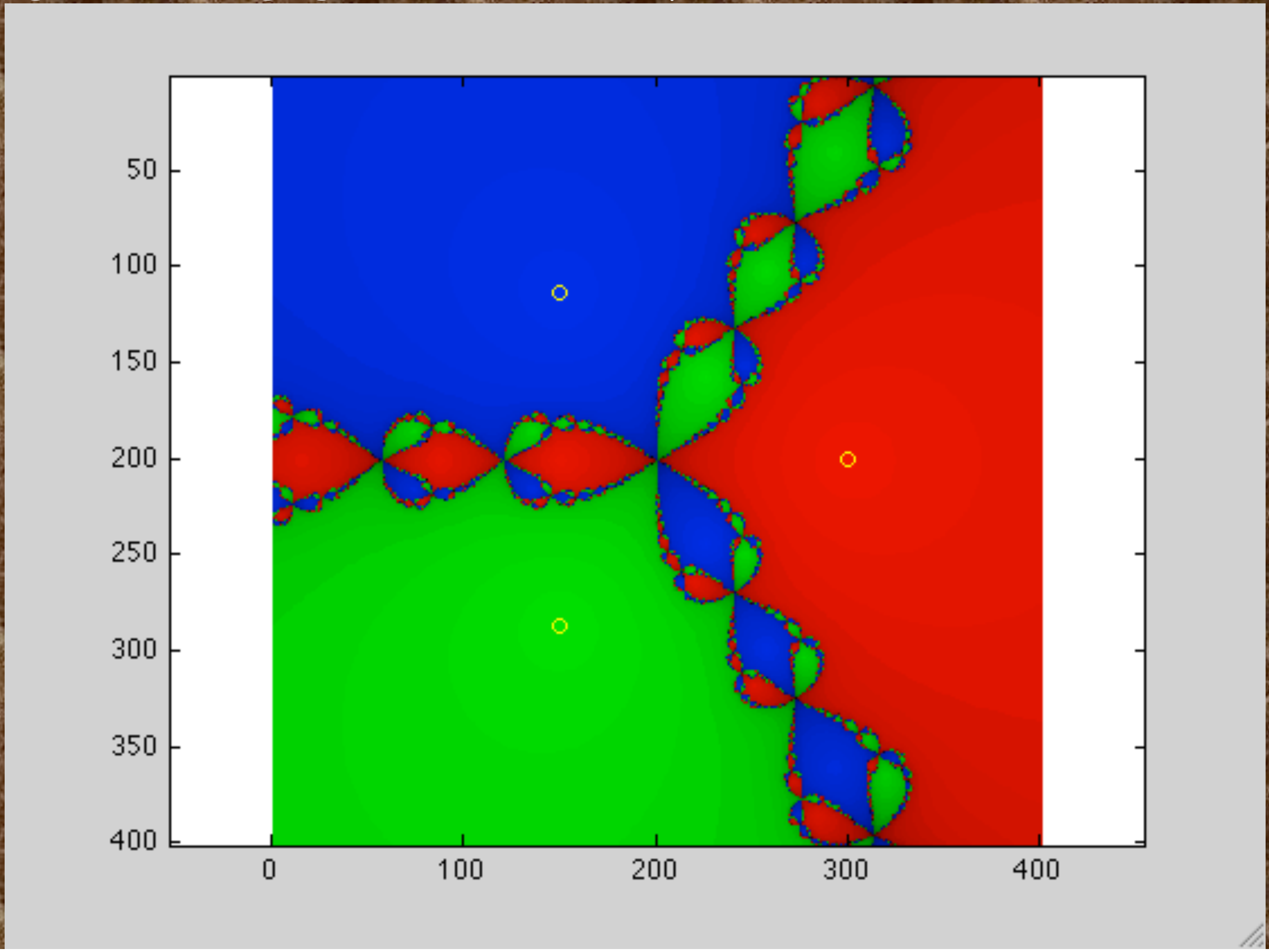


Here is what we get when we apply Newton's method and keep track of to where (color) and how fast (brightness/intensity) we converge..

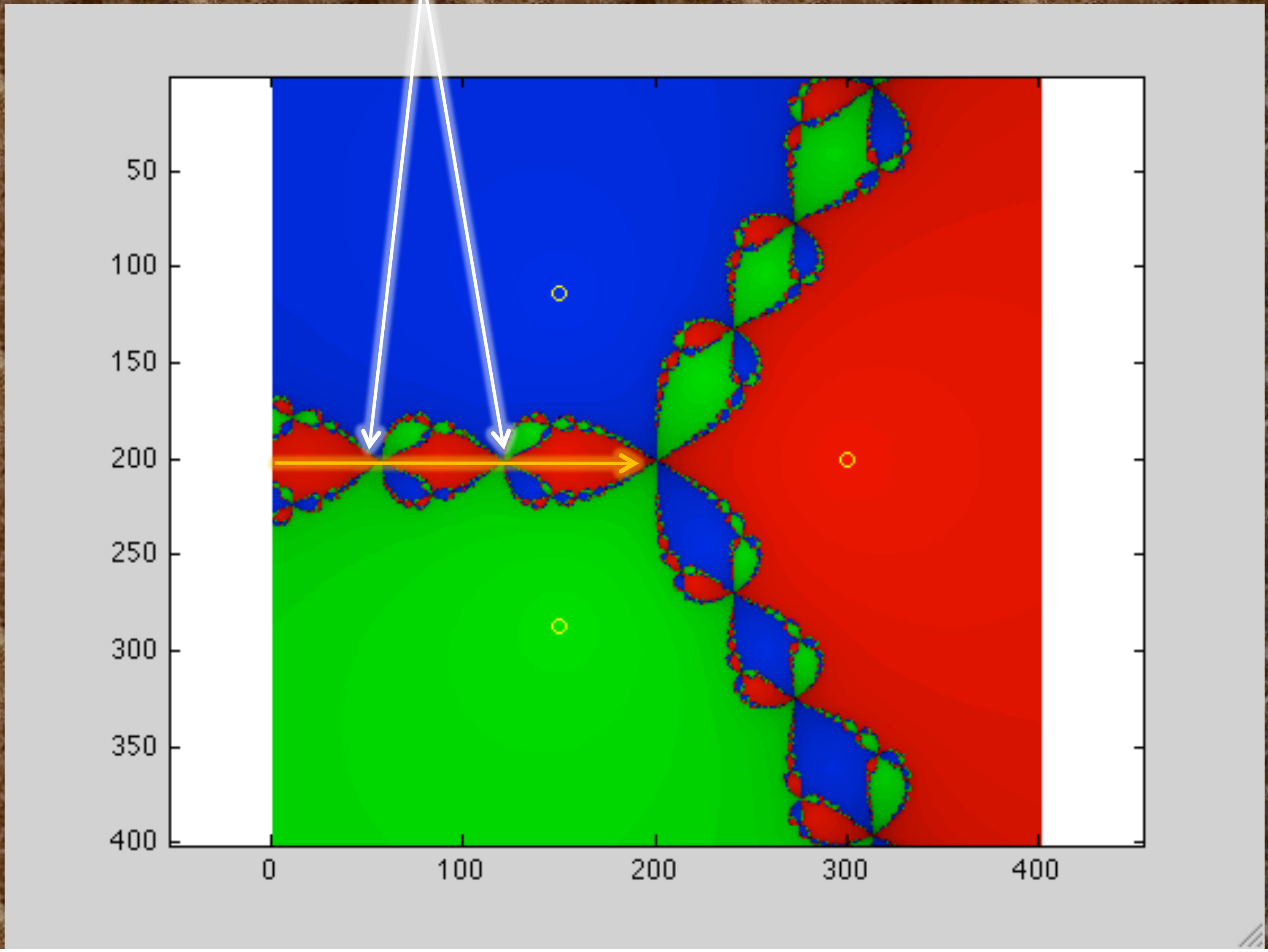


Notice that for most starting values the convergence is well behaved.

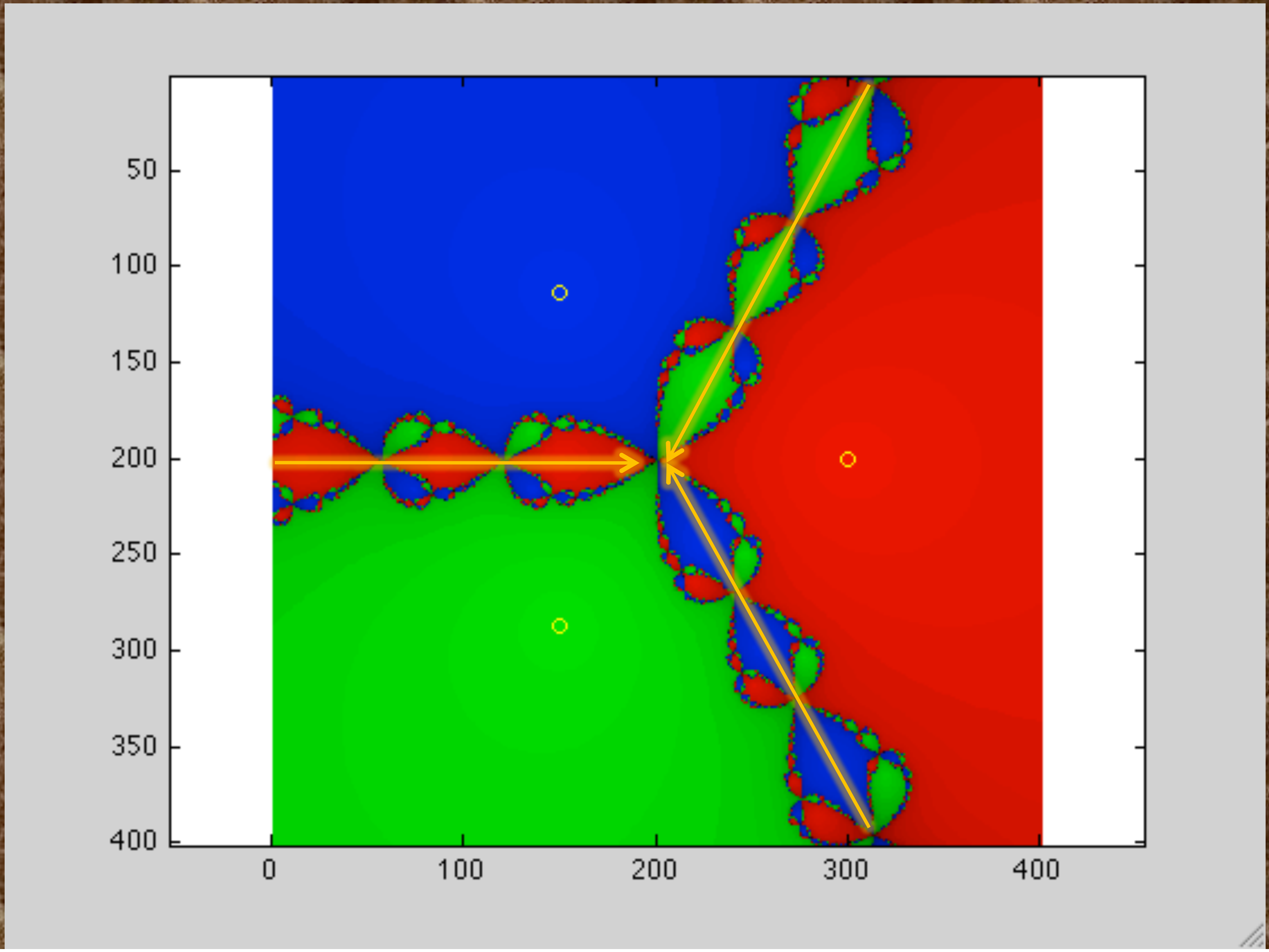
(the roots are arranged every 120°, convergence is relatively rapid , and the regions converging to each root are pie like sectors of 120° "width".)



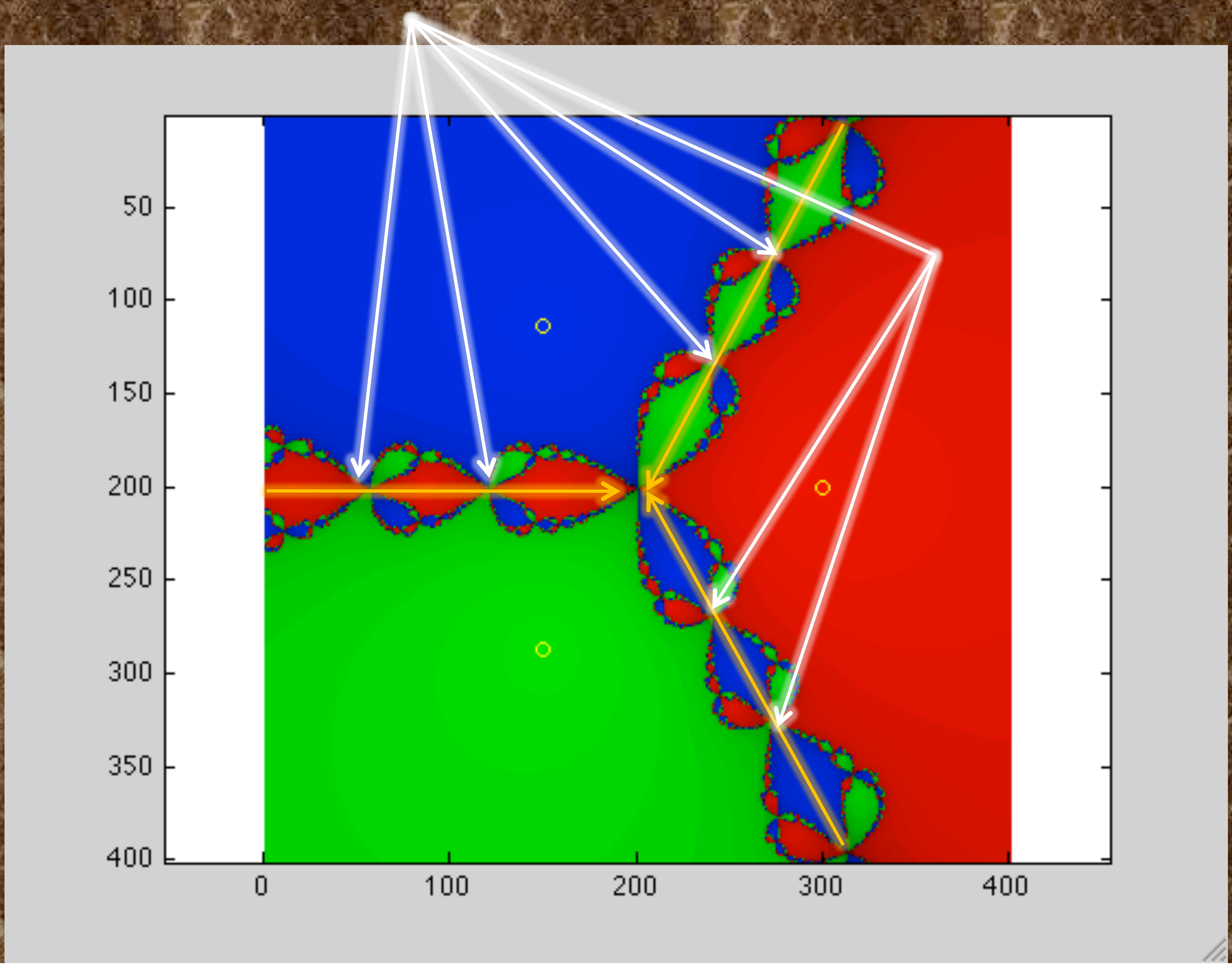
But there are some problematic places. Remember we had some problems along the negative real axis for the real case (where f-=0).



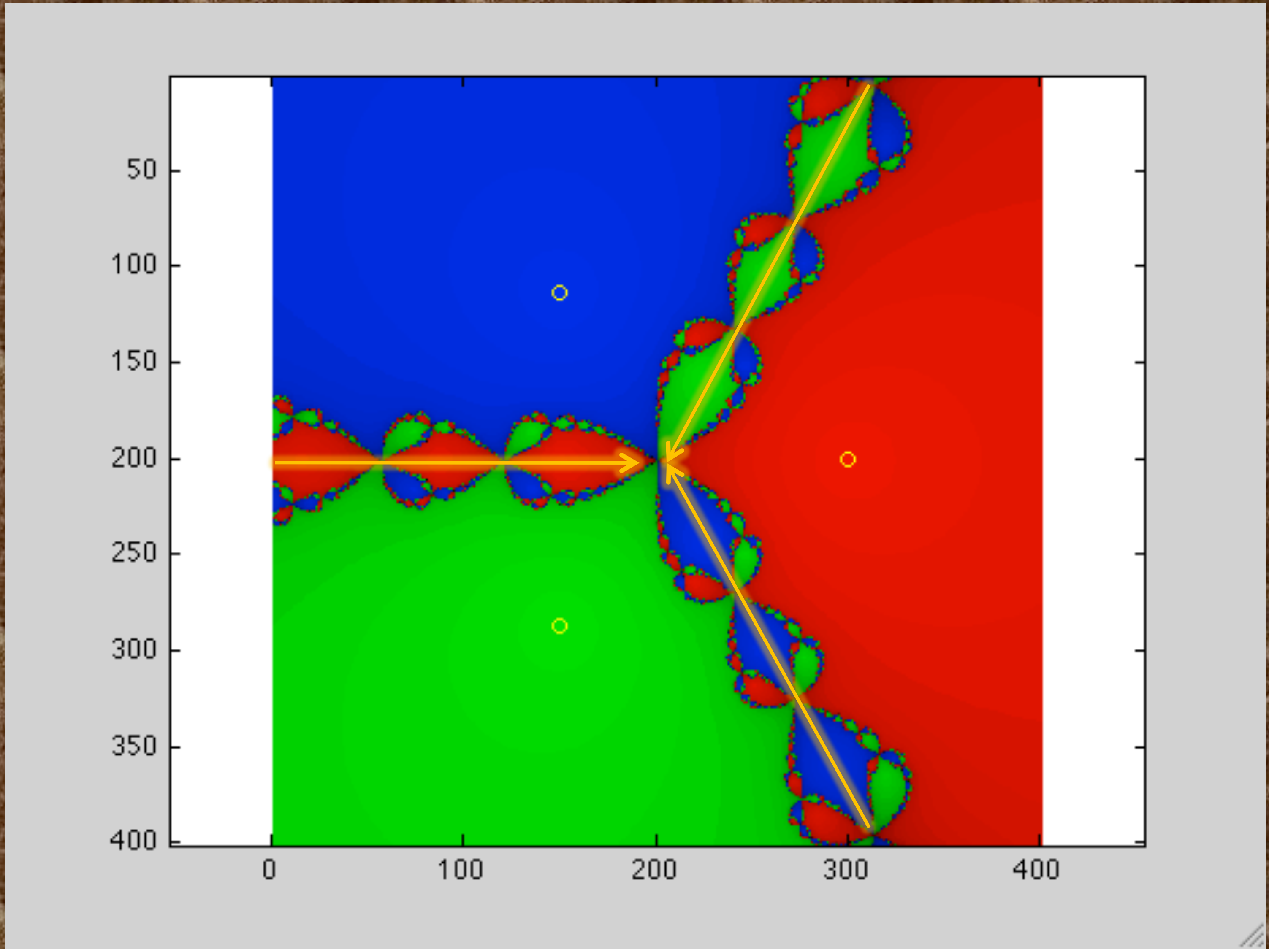
The starting values along the lines pointing towards the roots from the “other half plane” all have similar problems.



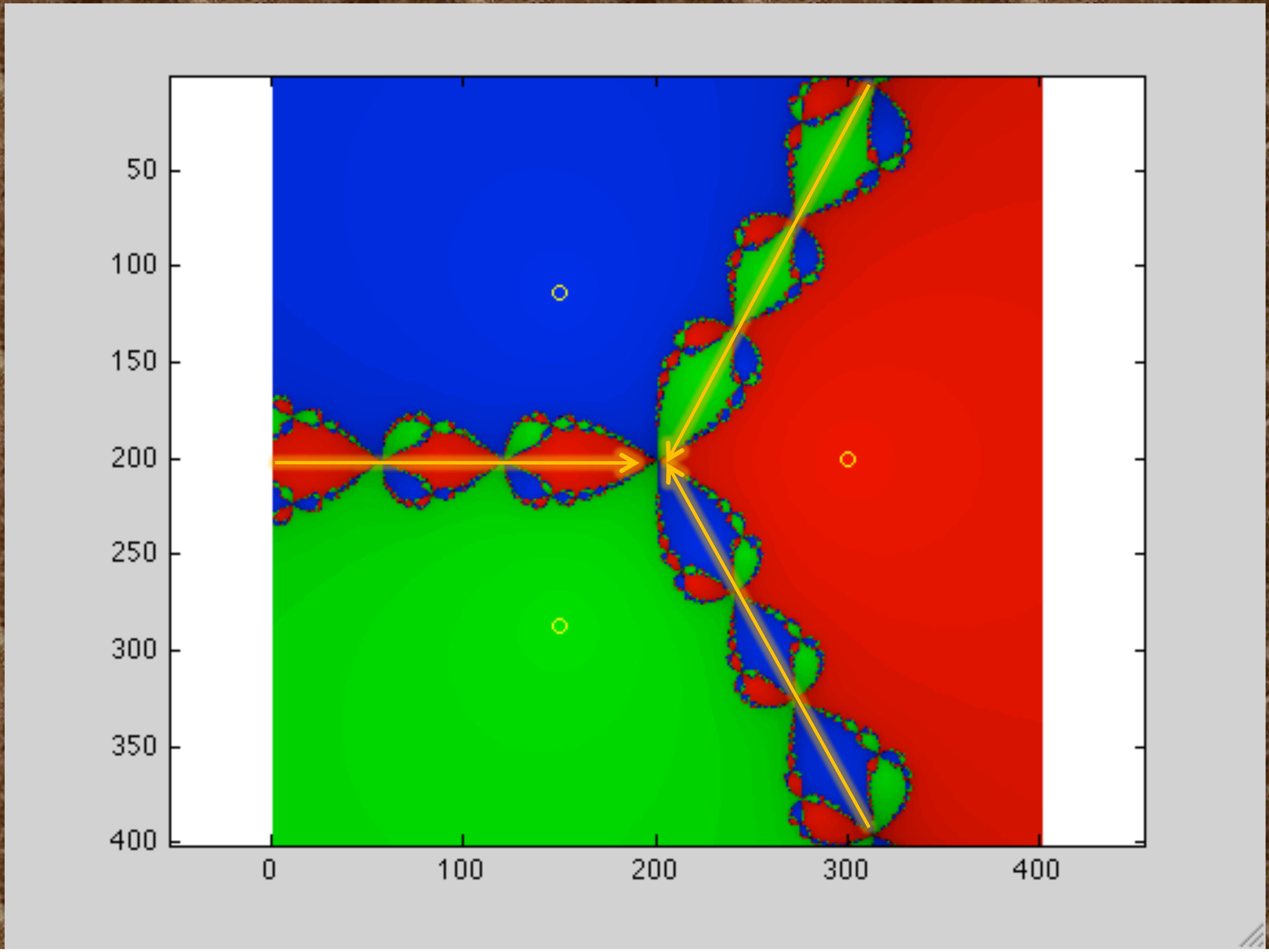
They are part of a structure of places with problems (same problem, they iterate to, or close to, z=0, where the slope is zero).



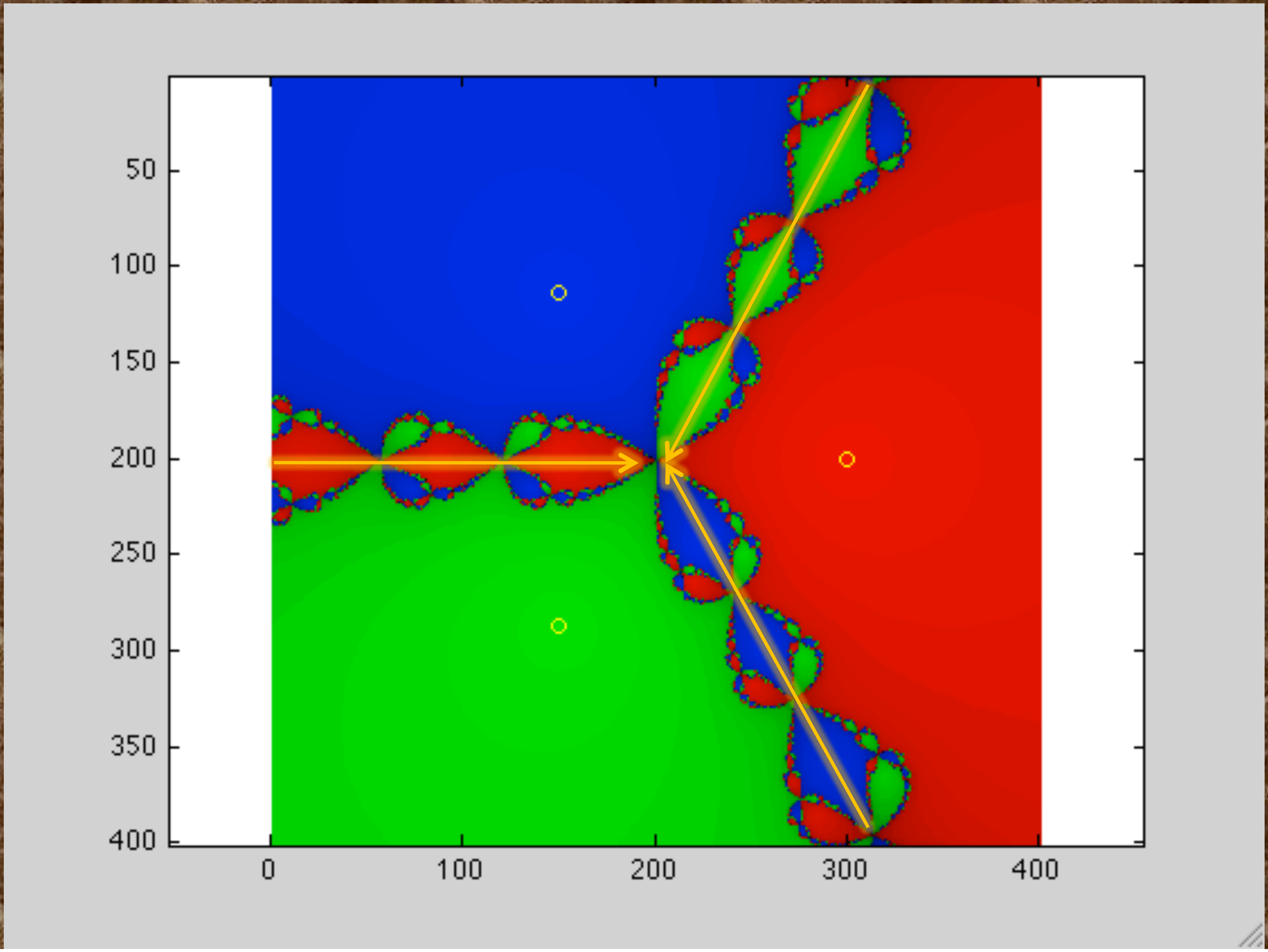
In addition, the division between regions of two different colors contains a region of the third color (except the points of no or slow convergence).

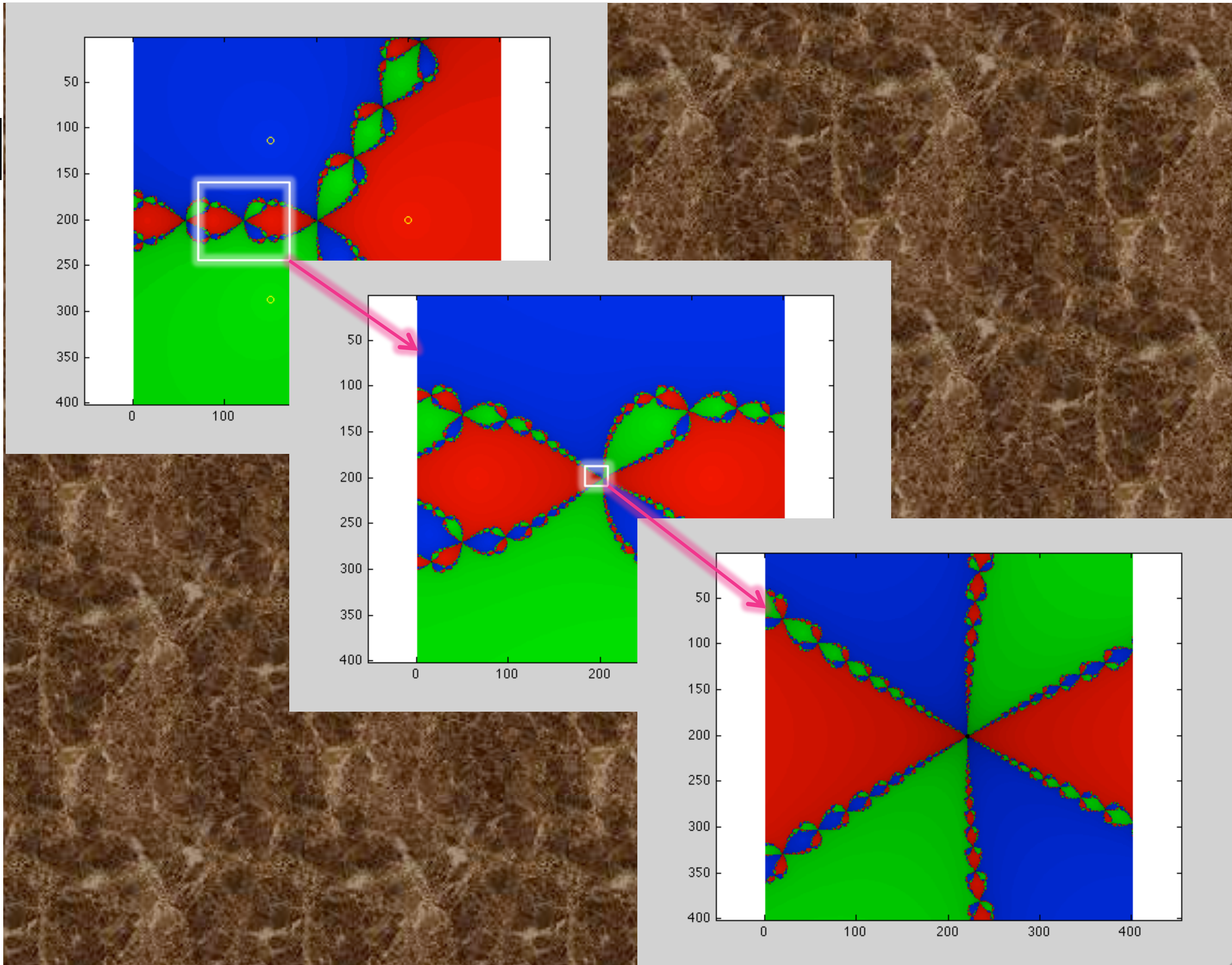


If one watched the path from these starting points, they would get near the problem points on the 3 lines dividing the 3 regions.



This color pattern behavior continues on all "boundaries" between colors, even as one zooms in on the boundaries.





etc.

In most geophysical inversions we do not look carefully at the convergence properties of the system.